Comparative Genomic and Metabolomic Analysis of Twelve Strains of Pseudoalteromonas luteoviolacea

Månsson, Maria; Vynne, Nikolaj Grønnegaard; Klitgaard, Andreas; Melchiorsen, Jette; Dorrestein, Pieter C.; Gram, Lone

Publication date:
2014

Document Version
Publisher's PDF, also known as Version of record

DTU Systems Biology
Department of Systems Biology

Comparative Genomic and Metabolomic Analysis of Twelve Strains of *Pseudoalteromonas luteoviolacea*

Maria Månsson[1], Nikolaj G. Vynne[1], Andreas Klitgaard[1], Jette Melchiorsen[1], Pieter C. Dorrestein[2], and Lone Gram[1]

[1]Department of Systems Biology, Technical University of Denmark; [2]Skaggs School of Pharmacy and Pharmaceutical Sciences, Departments of Pharmacology, Chemistry and Biochemistry, University of California San Diego

With the availability of full genome sequences, it has become apparent that the biosynthetic potential of many microorganisms is much larger than hitherto thought. Mining for new chemical diversity can be done 'upstream', directly at the genome level, or 'downstream', at the metabolite level. Here, we describe the biosynthetic capabilities of a marine bacterial species, *Pseudoalteromonas luteoviolacea* – a prolific producer of secondary metabolites.

Full genome sequences are now available for routine analysis, and in parallel, mass spectrometers are pushing the limits for sensitivity and accuray, generating high quality metabolite data at a high speed. Thus, the challenge is no longer to generate the data that describes secondary metabolite production in microorganisms, but to manually extract important elements of complex data sets and create a meaningful link between the observed chemistry, detected bioactivity, and predicted pathways. The bacterium, *Pseudoalteromonas luteoviolacea* is a prolific producer of secondary metabolites.[1] The genomes of 12 strains of *P. luteoviolacea* were sequenced and compared with regards to the presence of secondary metabolite clusters.

The average genome size of all 12 strains were ~6 Mb with 9-17% of the genome allocated to secondary metabolism, corresponding to 6-20 PKS/NRPS clusters predicted by antiSMASH . Overall, the accessory or variable genes made up more than 65% of the total pan-genome **(Fig. 1).** However, looking only at predicted biosynthetic genes they add up to almost 90% of the pan-genome, underlining the bio-synthetic heterogeneity of the species

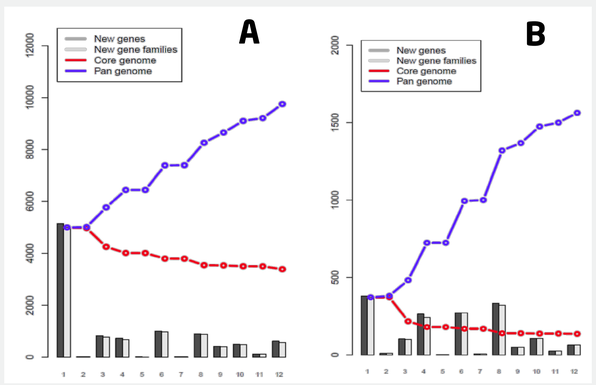

Figure 1. Pan- and core-genome plots based on the 12 analyzed genomes. The pan-genome curve (blue) connects the cumultative number of gene families present. The core-genome curve (red) connects the conserved number of gene families. The bars show the number of novel gene families identified in each genome. **A)** Plot based on all genes families; **B)** Gene families predicted by antiSMASH to be involved in secondary metabolism.

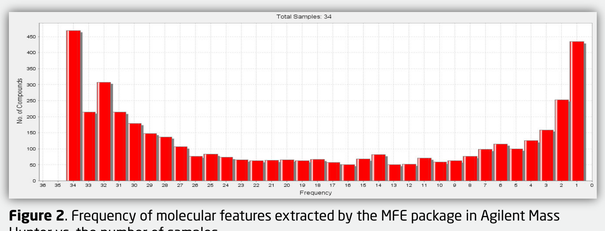

Figure 2. Frequency of molecular features extracted by the MFE package in Agilent Mass Hunter vs. the number of samples.

Matching our observations from the genome to the potential compounds produced by the encoding proteins, we used unbiased feature detection with recursive analysis on our LC-MS data from all the strains **(Fig. 2).**

Here, the great diversity within the species was confirmed, as only about 12% of the features were found in all strains (excluding features from media blank). Also, 15% of all features were only found in a single strain (unique features).

The number of predicted biosynthetic clusters and detected molecular features by far exceed the number of compounds known from this species. In order access those and at the same time gain information on the structural relationships, we have used molecular networking[2,3] on LC-MS/MS data (**Fig. 3**) .

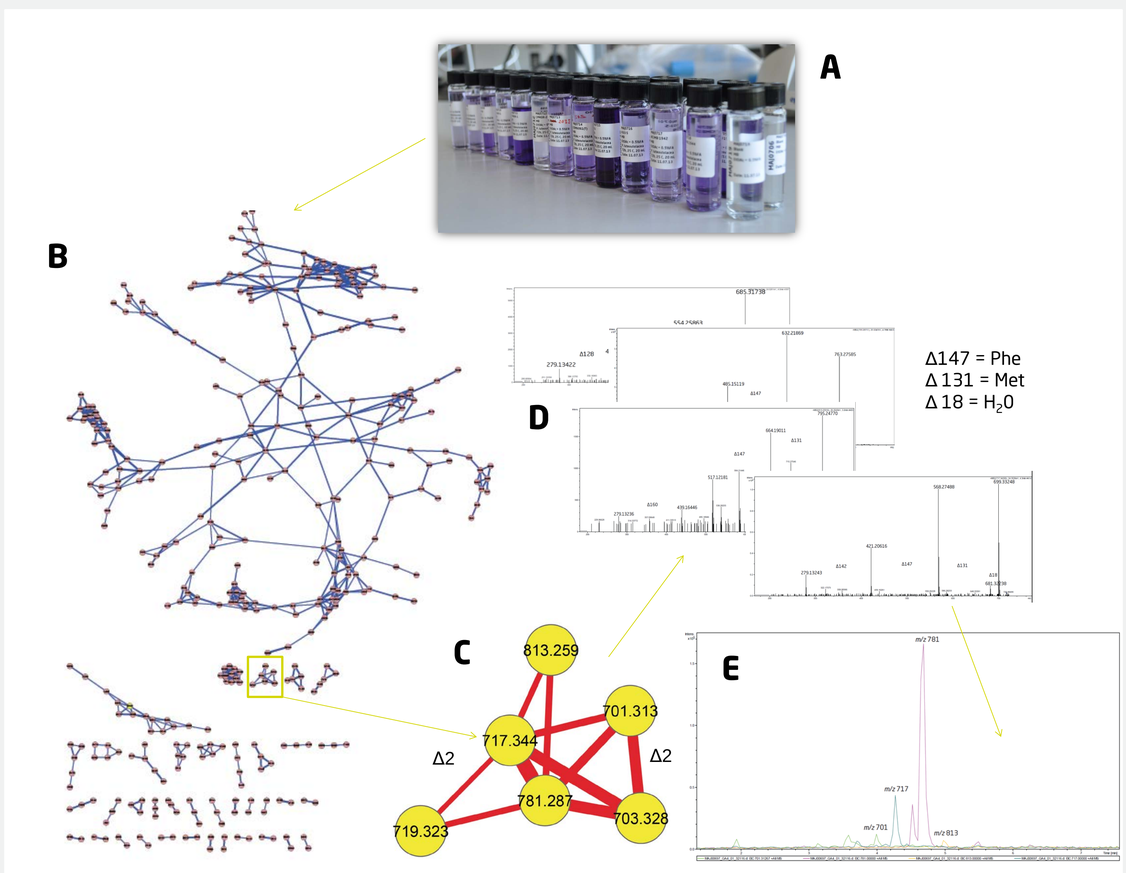

Figure 3. A) Extracts of 12 strains of *P. luteoviolacea*; **B)** Molecular network (based on Bruker LC-MS data) of *P. luteoviolacea* pan-metabolome; **C)** Unique molecular family based on halogenated peptides; **D)** LC-MS/MS spectra of compounds associated with molecular family, identifying characteristic fragments of phenylalanine, methionine, and loss of water; **E)** Deconvoluted LC-MS trace, high-lighting a series of analogues suitable for purification.

Conclusions:

By using a combined genomic and metabolomic approach to mine the chemical diversity within a single species, it is possible to quickly identify strains that hold novel chemistry and link the compounds to their biosynthetic pathway.

References: 1) Vynne *et al.* (**2012**), *Mar. Drugs.*, 10:1729-40; Yang *et al.* (**2013**), *J. Nat. Prod.*, 76(9):1686-99; 3) Watrous *et al.* (**2012**), *PNAS*, 109:E1743-52.

Funding: Danish Council for Independent Research in Technology and Production Sciences with Sapere Aude grant # 11 62 62 *"Ecology-Driven Drug Discovery: A Novel Approach for Enhancing Chemical Diversity from Marine Bacteria"* and FøSU grant # 2101-07-0012 *"Discovery of Bioactive Marine Bacteria and Natural Products and Their Use to Promote Human Health and Safety"*. Agilent Thought Leader Award for LC-QTOF data and software.

Acknowledgements: **Laura Sanchez**, **Jeramie Watrous**, and **Don Nguyen** are greatly acknowledged for training in molecular networking.

DTU Systems Biology
Department of Systems Biology
DTU Systems Biology
Department of Systems Biology
DTU Systems Biology
Department of Systems Biology
DTU Systems Biology
Department of Systems Biology